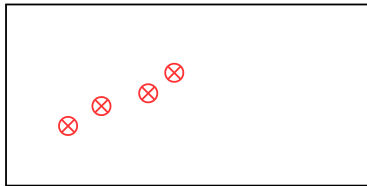


## Observed Spatial Dynamics

$(x_{\text{center}}, y_{\text{center}}, v_x, v_y, a_x, a_y)$

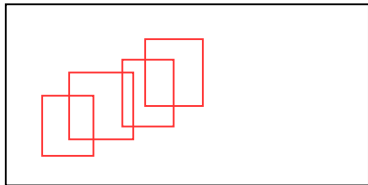


GRU-1  
Encoder

$h_{p,T}$

## Observed Dimensional Attributes

$(w, h, \Delta w, \Delta h)$



GRU-2  
Encoder

$h_{d,T}$

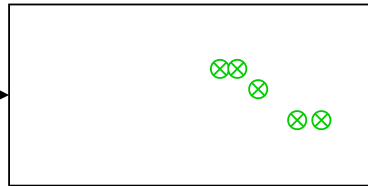
Concatenation



$h_{\text{combined}}$

## Predicted Spatial Dynamics

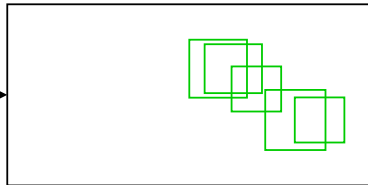
$(x'_{\text{center}}, y'_{\text{center}}, v'_x, v'_y, a'_x, a'_y)$



GRU-3  
Decoder

## Predicted Dimensional Attributes

$(w', h', \Delta w', \Delta h')$



GRU-4  
Decoder